Homework 1

jiaying Ning

9/23/2020

**Problem 1**

load packages

import data

1. Using the entire sample, provide descriptive statistics for all variables of interest.
2. Total sample size for each variable (N); Mean/SD, Median/IQR, Min and Max for continuous variables; Frequency/Percentages for categorical variables; Number of missing values for each variable. (8p)
3. Generate a his togram for the Ig-M values and comment on its shape. (2p)

since the only continuous variable is Antibody\_Igm. We will be taking the descriptive statistics measure for this continuous variable.

below are its overall summary and desctiptive value table

#overall summary  
summary(Antibodies)

## Subject AgeCategory Antibody\_IgM Smell   
## Min. : 1.0 18-30:318 Min. :0.0480 Altered :1047   
## 1st Qu.: 632.5 31-50:810 1st Qu.:0.0690 Normal : 410   
## Median :1373.0 51+ :363 Median :0.0915 Unanswered/Others: 34   
## Mean :1413.8 Mean :0.1239   
## 3rd Qu.:2190.5 3rd Qu.:0.1290   
## Max. :2917.0 Max. :1.0475   
## NA's :1224   
## Gender   
## Female:981   
## Male :510   
##   
##   
##   
##   
##

#print out the class table for each variable in the dataset  
t(t(sapply(Antibodies, class)))

## [,1]   
## Subject "integer"  
## AgeCategory "factor"   
## Antibody\_IgM "numeric"  
## Smell "factor"   
## Gender "factor"

#print out the descriptive statistics  
descriptivestats = summary(na.omit(Antibodies$Antibody\_IgM))  
descriptivestats

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0480 0.0690 0.0915 0.1239 0.1290 1.0475

IQR(na.omit(Antibodies$Antibody\_IgM))

## [1] 0.06

sd(na.omit(Antibodies$Antibody\_IgM))

## [1] 0.1104458

From the dataset, I observe that only meaningful variable that is numeric is “Antibody\_IgM”. Variable “AgeCategory”,“Smell”,“Gender” are all factor.

For continuous variable: Antibody\_Igm has means of 0.1238839 with standard deviation of 0.1104458, median of 0.0915 and IQR of 0.06.Its min is 0.048 and max is 1.0475

For categorical variable the frequency and proportion of agegroup is:

#frequency for categorical variable  
##frequency of age category  
table(Antibodies$AgeCategory)

##   
## 18-30 31-50 51+   
## 318 810 363

prop.table(table(Antibodies$AgeCategory))

##   
## 18-30 31-50 51+   
## 0.2132797 0.5432596 0.2434608

the frequency and proportion of subject’s smell status is:

##frequency of smell  
table(Antibodies$Smell)

##   
## Altered Normal Unanswered/Others   
## 1047 410 34

prop.table(table(Antibodies$Smell))

##   
## Altered Normal Unanswered/Others   
## 0.70221328 0.27498323 0.02280349

the frequency and proportion of smell is:

##frequency of Gender  
table(Antibodies$Gender)

##   
## Female Male   
## 981 510

prop.table(table(Antibodies$Gender))

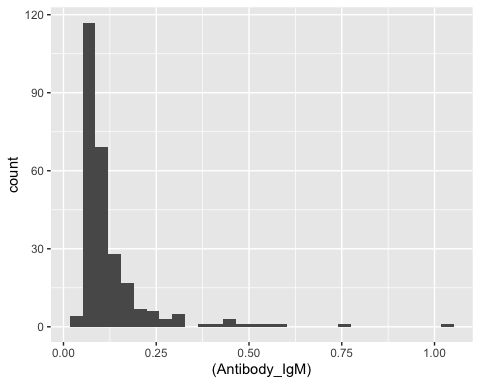
##   
## Female Male   
## 0.6579477 0.3420523

From the summary table we know that there are 1224 NA values in Antibody\_Igm variable. 34 unanswered/other in smell category.

ggplot(data=Antibodies,mapping=aes((Antibody\_IgM))) + geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 1224 rows containing non-finite values (stat\_bin).

 From the graph showing above, we can see that the distribution is skewing to the right. Most people have low antibody\_igm.

**problem 2**

1. Provide descriptive statistics for all variables, stratified by smell category.
2. Provide descriptive statistics for all the other variables by the two smell categories (normal vs altered) and present them in a tabular form (see example below). Briefly comment on the differences observed between the two groups. (6p) Hint: tableby() in R can easily generate this, but feel free to create your own table in Word.

library(dplyr)  
library(furniture)  
  
table1(Antibodies,AgeCategory,Subject,Antibody\_IgM,Gender,splitby = "Smell", row\_wise = TRUE)

##   
## ────────────────────────────────────────────────────────────  
## Smell   
## Altered Normal Unanswered/Others  
## n = 178 n = 81 n = 8   
## AgeCategory   
## 18-30 36 (67.9%) 13 (24.5%) 4 (7.5%)   
## 31-50 101 (68.2%) 43 (29.1%) 4 (2.7%)   
## 51+ 41 (62.1%) 25 (37.9%) 0 (0%)   
## Subject   
## 585.9 (517.6) 622.9 (671.0) 722.4 (622.8)   
## Antibody\_IgM   
## 0.1 (0.1) 0.1 (0.1) 0.1 (0.1)   
## Gender   
## Female 120 (69%) 47 (27%) 7 (4%)   
## Male 58 (62.4%) 34 (36.6%) 1 (1.1%)   
## ────────────────────────────────────────────────────────────

Antibodies\_clear = filter(Antibodies, Smell != "Unanswered/Others")  
table1(Antibodies\_clear,AgeCategory,Subject,Antibody\_IgM,Gender,splitby = "Smell", row\_wise = TRUE)

##   
## ──────────────────────────────────────────  
## Smell   
## Altered Normal   
## n = 178 n = 81   
## AgeCategory   
## 18-30 36 (73.5%) 13 (26.5%)   
## 31-50 101 (70.1%) 43 (29.9%)   
## 51+ 41 (62.1%) 25 (37.9%)   
## Subject   
## 585.9 (517.6) 622.9 (671.0)  
## Antibody\_IgM   
## 0.1 (0.1) 0.1 (0.1)   
## Gender   
## Female 120 (71.9%) 47 (28.1%)   
## Male 58 (63%) 34 (37%)   
## ──────────────────────────────────────────

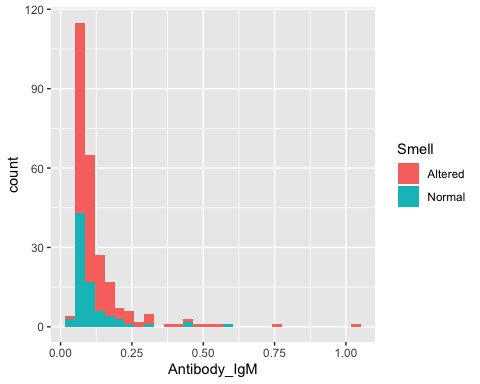
Overall, more people reported having their smell altered than not. This difference in proportion is true among all age group and gender.

1. Use the Ig-M variable to generate side-by-side histograms and boxplots by smell categories

ggplot(Antibodies\_clear,aes(x=Antibody\_IgM,fill=Smell))+geom\_histogram()

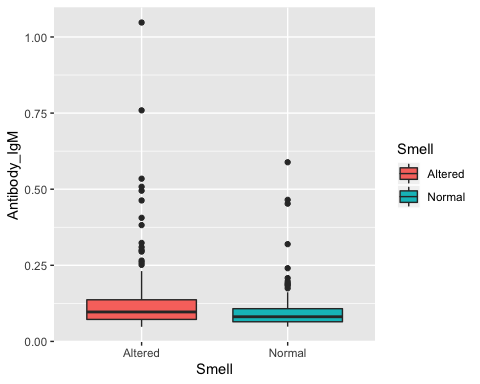
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 1198 rows containing non-finite values (stat\_bin).



ggplot(Antibodies\_clear, aes(x=Smell, y=Antibody\_IgM, fill=Smell)) +   
 geom\_boxplot()

## Warning: Removed 1198 rows containing non-finite values (stat\_boxplot).



(normal vs altered). Make sure you label your figures appropriately and briefly discuss the trends observed. (4p)

Btoh histogram skewed to the right, with most participants having relatively lower IGM value. However, the variation for altered-smell participants seems to be larger than normal participants. The average of IGM is also higher than altered-smell group. Further more, the altered-smell group seem to have more extremed IGM value than normal group.

##problem 2

In this article, author discussed a study done by Professor Shoveller who suggested that cats that eats only one meals a day might feel more satisfied and have healthier body than cats who eats several small meals a day. To present this fact, the author compares this result to another more widely accepted theory, which is that “eating small meals several days is better for cats”. Then the author shortly explained how the study set up and lay out the conclusion of the study.

Summary: The main goal of the study is to examine how feeding frequency impact cats’ mental and physical state. Total of 8 cats with similar age are included in the study, all cats are assigned with the same food. The study used a 2X3 replicated incomplete Latin square design. The study does not specify whether cats are randomly assigned to treatment. Some potential bias might include: 1. Even though precise measurement for health and behavior is recorded for each cat, the sample size might still be too small to be generalized to a larger cat population 2. Different feeder prepares different type of food for their cat, and the quality of food can have strong impact on a cats’ mental and physical state. Therefore, since all cats are assigned with the same canned food in the study, it is not clear whether the food choice can be a potential confounding variable.

Journal/Newspaper Reference: Leggate, J. (2020, September 24). You might be feeding your cats wrong, according to a new study. Fox News. <https://www.foxnews.com/lifestyle/feeding-cats-wrong-new-study> Original Study reference: Camara A, Verbrugghe A, Cargo-Froom C, Hogan K, DeVries TJ, et al. (2020) The daytime feeding frequency affects appetite-regulating hormones, amino acids, physical activity, and respiratory quotient, but not energy expenditure, in adult cats fed regimens for 21 days. PLOS ONE 15(9): e0238522.